What is Claimed:

1. A method for dynamically updating information for publication comprising:

a) extracting from received information a set of characterizing features which characterize the received information;

b) grouping together received information having common characterizing features into a number of clusters; and

c) using the information obtained in the grouping step to publish information contained in a cluster based on a customer request for information.

2. The method of claim 1 wherein the received information comprises a combination of one or more of text data, image data, or video data.

3. The method of claim 1 wherein said received information comprises multiple features of a given type and wherein the multiple features are ranked in importance as the features are extracted.

4. The method of claim 3 wherein a cluster includes a summarization of cluster features and additionally comprising comparing the features that summarize newly received information with features summarized in a cluster by taking an inner product of the features common to the newly received information and the features that summarize said cluster and combining the newly received information with a cluster if the inner product exceeds a threshold.

5. The method of claim 1 wherein the top K features of rank of a newly received item of information are compared with the top K features of a cluster to determine if said information is added to a cluster.

6. The method of claim 5 wherein each feature has a relevancy factor by which the feature is scaled and additionally determining if a cluster and the newly received information have at least L common features having non-zero relevancy factors before adding the received information into a cluster.

7. The method of claim 1 additionally comprising grouping together clusters having a common characteristics to produce a neighborhood of clusters which are all published in response to a customer request.

8. The method of claim 7 wherein the received information is a text containing document and a relevancy of a neighborhood is used to determine whether to publish documents in a neighborhood to a customer.

9. The method of claim 8 wherein the relevancy varies depending on how long the document has been assigned to the neighborhood.

10. The method of claim 8 wherein the relevancy varies with information contained in the request for information.

11. The method of claim 7 wherein an item of received information may be grouped into more than one cluster but published with only one neighborhood.

12. The method of claim 11 additionally comprising maintaining a null neighborhood and adding received information to the null neighborhood when said information is initially received.

13. The method of claim 11 additionally comprising maintaining a null neighborhood and adding received information to the null neighborhood when contents of a neighborhood change due to a reconstituting of said neighborhood.

14. The method of claim 11 additionally comprising maintaining a null neighborhood and adding received information to the null neighborhood when a neighborhood to which the received information becomes non-relevant.

15. A process for evaluating documents comprising:

a) evaluating multiple documents containing text data for subsequent publication by extracting K tokens having a highest token relevance factor based on the frequency of token occurrence within the document;

b) grouping together documents having a commonality in said text data that is greater than a threshold to provide a number of document clusters of said documents; said grouping performed by:

i) comparing the K tokens from a candidate document with a document cluster characterizing set of tokens;

ii) adding a candidate document to a document cluster if the comparison indicates a sufficient degree of similarity between the candidate document and said document cluster; and

iii) updating a document cluster summarization that takes into account the added candidate document; and

c) publishing documents assigned to a specified document cluster or document clusters based upon a request.

16. The process of claim 15 wherein documents have document categories and evaluating the token relevance factor comprises determining a category frequency of tokens within a document category and assigning a relevance factor to said token based on said category frequency.

17. The process of claim 16 wherein tokens are assigned a relevance factor based on a position of a token within the document.

18. The process of claim 15 wherein if said candidate document is not sufficiently similar to a cluster it forms the basis of its own new cluster.

19 The process of claim 15 wherein the token relevance factor is determined from a relation $\exp(-a*p_{0i})*N_i*R_i$, where a is the decay rate of token relevance as a function of the distance from the beginning of the text of a document D, p_{0i} is the position at which token i first appears in the text, N_i is the number of occurrences of token i and R_i is the log of the

inverse document frequency of token i in the category of documents to which the document D belongs.

20. The process of claim 15 wherein clusters of documents are clustered to form neighborhoods of documents to which documents are assigned.

21. The process of claim 16 wherein the neighborhoods are assigned a neighborhood relevancy factor which varies with time, said neighborhood relevancy factor used to determine to whom a neighborhood is published.

22. The process of claim 21 wherein the neighborhood relevancy number also varies with a document relevancy factor of documents that make up the neighborhood.

23. The process of claim 22 wherein the document relevancy factor depends on the quality of the source of the document.

24. The process of claim 22 wherein the document relevancy factor depends on the location of the source and the location of a requestor.

25. The process of claim 21 wherein the neighborhood relevancy factor varies with the category of documents assigned to said neighborhood.

26. A system for evaluating documents comprising:

a) a preprocessor for receiving text documents from one of a plurality of document sources and evaluating text data contained in each received document for determining suitability of the document for subsequent publication based on a request; said preprocessor grouping together documents having a commonality greater than a threshold to provide a number of clusters of said documents; and

b) a web server having access to the cluster data from the preprocessor for making available to a requester documents contained within a cluster based a comparison between

a request from the requester and a summarization of text contained within documents of a specified cluster or clusters.

27. The system of claim 26 wherein the preprocessor groups together clusters into a neighborhood of clusters and further wherein documents within a neighborhood are made available to a requester.

28. The system of claim 26 wherein a cluster of documents is removed from publication by the web server based on a cluster relevancy of the entire cluster.

29. A computer readable medium containing instructions for dynamically updating information for publication comprising instructions for:

a) extracting from received information a set of characterizing features which characterize the received information;

b) grouping together received information having common characterizing features into a number of clusters; and

c) using the information obtained in the grouping step to publish all information contained in a cluster based on a customer request for information.

30. The computer readable medium of claim 29 wherein the received information comprises a combination of one or more of text data, image data, or video data.

31. The computer readable medium of claim 29 wherein said received information comprises multiple features of a given type and wherein the multiple features are ranked in importance as the features are extracted.

32. The computer readable medium of claim 31 wherein a cluster includes a summarization of cluster features and additionally comprising comparing the features that summarize newly received information with features summarized in a cluster by taking an inner product of the features common to the newly received information and the features

that summarize said cluster and combining the newly received information with a cluster if the inner product exceeds a threshold.

33. The computer readable medium of claim 29 wherein the top K features of rank of a newly received item of information are compared with the top K features of a cluster to determine if said information is added to a cluster.

34. The computer readable medium of claim 33 wherein each feature has a relevancy factor by which the feature is scaled and additionally determining if a cluster and the newly received information have at least L common features having non-zero relevancy factors before adding the received information into a cluster.

35. The computer readable medium of claim 29 comprising an additional step of grouping together clusters having a common characteristics to produce a neighborhood of clusters which are all published in response to a customer request.

36. The computer readable medium of claim 35 wherein a relevancy of a neighborhood is used to determine whether to publish documents in a neighborhood to a customer.

37. The computer readable medium of claim 36 wherein the relevancy varies with how long the document has been in the neighborhood.

38. The computer readable medium of claim 36 wherein the relevancy varies with information contained in a request for information.

39. The computer readable medium of claim 35 additionally comprising maintaining a null neighborhood and adding received information to the null neighborhood when said information is initially received.

40. The computer readable medium of claim 35 additionally comprising maintaining a null neighborhood and adding received information to the null neighborhood when contents of a neighborhood change due to a reconstituting of said neighborhood.

41. The computer readable medium of claim 35 additionally comprising maintaining a null neighborhood and adding received information to the null neighborhood when a neighborhood to which the received information becomes non-relevant.